

The Delphion Integrated View: INPADOC Record

Get Now: ☒ PDF | [File History](#) | [Other choices](#)

Tools: [Add to Work File](#) | [Create new Work File](#)

View: Jump to: [Top](#) [Go to: Derwent](#)

[Ema](#)

🔍 Title: **CN1191349A: Interconnection network of which bandwidth is exten
method for transferring data in the same**

🔍 Derwent Title: Multi-processor interconnection system - forwards message using data network whose interconnection is performed by same topology which is used for interconnection of control network ([Derwent Record](#))

🔍 Country: CN China

🔍 Kind: A Unexamined APPLIC. open to Public inspection 

🔍 Inventor: KAB-YOUNG KIM; Republic of Korea

🔍 Assignee: SAMSUNG ELECTRONICS CO., LTD. Republic of Korea
[News, Profiles, Stocks and More about this company](#)

🔍 Published / Filed: 1998-08-26 / 1998-01-26

🔍 Application Number: CN1998000103706


🔍 IPC Code: Advanced: [G06F 15/16](#); [G06F 15/177](#); [H04L 12/56](#);
Core: more...
IPC-7: [G06F 13/00](#); [G06F 15/173](#);

🔍 ECLA Code: H04L12/56C5;

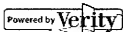
🔍 Priority Number: 1997-01-27 [KR1997000002281](#)

🔍 INPADOC Legal Status: None [Get Now: Family Legal Status Report](#)

🔍 Family:

PDF	Publication	Pub. Date	Filed	Title
	US6304568	2001-10-16	1998-01-26	Interconnection network extendable band method of transferring data therein
<input checked="" type="checkbox"/>	KR0259276B1	2000-06-15	1997-01-27	INTERCONNECTION NETWORK HAVIN EXTENDABLE BANDWIDTH
<input checked="" type="checkbox"/>	JP10222480A2	1998-08-21	1998-01-27	MUTUAL CONNECTION NETWORK AN TRANSFER METHOD
<input checked="" type="checkbox"/>	CN1191349A	1998-08-26	1998-01-26	Interconnection network of which bandwi extendable and method for transferring d: same
<input checked="" type="checkbox"/>	CN1108571C	2003-05-14	1998-01-26	Bandwidth extendable interconnection ne method for transferring data in the same
5 family members shown above				

🔍 Other Abstract Info: [DERABS G98-511417](#)



[Nominate this for the Gallery...](#)

Copyright © 1997-2008 The Thom

[Subscriptions](#) | [Web Seminars](#) | [Privacy](#) | [Terms & Conditions](#) | [Site Map](#) | [Contact U](#)



[12] 发明专利申请公开说明书

[21] 申请号 98103706.2

[43] 公开日 1998 年 8 月 26 日

[11] 公开号 CN 1191349A

[22] 申请日 98.1.26

[30] 优先权

[32] 97.1.27 [33] KR[31]2281/97

[71] 申请人 三星电子株式会社

地址 韩国京畿道

[72] 发明人 金甲英

[74] 专利代理机构 中国国际贸易促进委员会专利商标

事务所

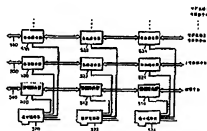
代理人 郑 迅

权利要求书 1 页 说明书 6 页 附图页数 7 页

[54] 发明名称 带宽可扩展的互连网络以及在其中传送数据的方法

[57] 摘要

具有分离的控制平面和数据平面且带宽可扩展的互连网络, 包括一个控制平面和一个或多个数据平面, 控制平面包括若干控制路由器和一条控制线, 控制路由器用于交换控制信息及生成设定报文传送通路的信息, 各个数据平面包括若干数据路由器和一条数据传送线, 每个数据路由器用于利用控制路由器生成的报文上的通路信息发送和接收报文, 数据传送线用于把数据路由器按和控制平面相同的拓扑结构连接在一起。



说明书

带宽可扩展的互连网络

以及在其中传送数据的方法

本发明涉及多计算机系统中的互连网络,尤其涉及其中包括控制平面和数据平面的并且该数据平面的带宽是可扩展的互连网络以及在该互连网络中传送数据的方法。

由于结构问题和物理限制,常规冯·诺伊曼计算机的性能是有限的。为了克服这样的限制把不同的计算机互连起来以允许并行处理。即,连接多个单处理机或者多个对称的多处理机(SMP)以并行处理并实现高性能计算机系统。采用这种多计算机并行处理方法的系统包括预定数量的节点和一个用于在节点之间传送报文的互连网络。

在这种情况下,系统的总性能取决于每个节点的性能以及连接各节点的该互连网络的性能。互连网络的性能是用节点间的带宽和等待时间确定的。从而,为了在短时间内从一个源节点向目的地节点发送大量的数据,互连网络必须提供宽的带宽和小的等待时间。尤其,例如决策支持系统(DSS)和多媒体系统的新领域需要在节点之间传送大量的数据。这些应用造成需要改进互连网络的性能。

图1表示多计算机系统中互连网络和节点的总结构。互连网络提供通过其节点可互相通信的通路。它们可以具有不同的结构,例如环、Banyan树、树、超立方体、网和环面。上述的结构各具有优点和缺点。常规互连网络的公共特性将参照网状结构说明。

图2表示具有普通二维网状结构的网络。在图2中,各个交叉点代表一个网络节点。节点包括计算部件(CU)和路由器。路由器和其它相邻节点连接并且当把互连网络和CU考虑成独立元素时构成部分互连网络。

当要在节点之间传送报文时必须首先建立通路。从线路交换方法、包交换方法、蛀洞(wormhole)路由方法以及虚拟开凿方法中选出的

一种适当方法可用来设定通路。通常，在设计路由器时确定通路设定方法。因为在确定互连网络的等待时间中通路的设定是很重要的，在选择设定通路的方法之前必须全面考虑各种因素，例如系统的最终使用、报文的长度和传送频率。

通常，一个路由器为了设定通路、管理互连网络以及检测网络错误需要和其它相邻的路由器交换特定目的的信息。在常规互连网络中，这些信息和普通的数据一起共享互连网络的相同链路。但是，在这种情况下不能有效地构造互连网络的链路。即，在常规的互连网络里，当必须经常传送大量的数据时，用于设定通路或管理互连网络的控制信号和要传送的数据共享网络的链路，从而在它们之间产生不可避免的干扰。从而，损害有效带宽和等待时间。另外，因为必须同时处理控制信号和传送数据，路由器的设计变得麻烦并且确定通路所需的时间变得更长。

而且，当在诸如 NCR 公司的 BYNET、Tandem 公司的服务器网络和 IBM 公司的 HPS 的常规互连网络中设计互连网络的路由器时，路由器之间的链路的带宽是确定的。这样，在为了用于下一代应用需要改进互连网络的带宽和等待时间的情况下，常规的互连网络不提供进行改进的适宜机制。

通常，为了增加带宽，必须采用高速半导体技术或者增加半导体芯片的引脚数量。但是，这些方法需要已经达到的半导体集成技术中的许多成果，并且在许多情况中需要它的最突出成果。并行地重叠多个互连网络是一种提高有效总带宽的替代办法。利用这种办法有可能增加可经互连网络同时传送的标准数据的吞吐量。但是，这不能得到可以通过提高路由器中的链路的带宽得到的各种好处。即，在采用二个重叠的互连网络的情况下，有可能同时传送二条报文，但是不可能减少传送一条报文所需的时间。

在图 3 中，对按照常规技术重叠互连网络的情况和增加链路的带宽的情况进行比较。假设要发送的报文的长度为 M 并且常规互连网络的带宽为 W 。报文传输时间为 M/W ，因为在二个独立的互连网络情况

下选择网络 1 和网络 2 中的一个。同时，在加倍带宽的情况下，有可能把报文传输时间减小到 $M/2W$ 。因此，重叠互连网络的一个重要原因实际上是通过冗余提供错误容限而不是扩大带宽。此外，因为常规互连网络通常不能提供加倍带宽的作用，为了扩大带宽，必须重新设计路由器。

下面将参照附图说明常规互连网络共有的结构问题。假定互连网络具有网状结构。图 4A 至 4C 表示根据常规技术的网状平面内的报文传送过程。在图 4A 至 4C 中，A、B 和 C 代表源节点，而 A'、B' 和 C' 代表对应的目的地节点。可以采用线路交换方法、包交换方法和蛀洞路由方法中的任一方法。首先，在图 4A 中，已把报文的通路设定为从 A 到 A' 并正传送数据。B 和 C 处于设定通路的处理之中。此时，因为 A 已经使用 B 和 C 所需使用的通路的一部分，由 B 传送的报文的标题和由 C 传送的报文的标题必须分别在 (1, 1) 节点和 (2, 1) 节点中待用，直至 A 完成报文的传送并且释放它所使用的通路。当 A 完成报文的传送时，在 (1, 1) 节点待用的 B 的报文标题和在 (2, 1) 节点待用的 C 的报文标题可以着手设定到达它们各自的目的地节点的通路。图 4B 表示这样的过程。如图 4C 中所示，B 和 C 完成设定通路并且可以沿着设定的通路向各自的目的地节点传送报文。从而，当一条报文的标题被别的报文阻塞时，在常规互连网络中，由于设定通路的处理，等待时间变长。因为其它的节点在源节点和目的地节点释放阻塞的通路之前不能设定通路。

本发明的一个目的是提供一种多计算机系统互连网络，在该系统中该互连网络里的带宽是可扩展的并且要求相同的等待时间而不管数据的长度及频率，其中通过在互连网络中附加地包括一个控制平面和一个数据平面来设定通路并且该数据网络的带宽是可扩展的。

本发明的另一个目的是提供一种在上述互连网络中传送数据的方法。

为了达到第一个目的，提供一种其带宽可扩展的互连网络，用于在构成一个多计算机系统的预定数量的处理机之间发送和接收报文，其

包括一个控制平面和一个或更多的数据平面，控制平面包括和处理机连接的用于交换控制信息及生成与为在处理机之间传送的报文设定通路有关的信息的若干控制路由器，以及一条用于把控制路由器按预定的拓扑结构连接在一起的控制线，每个数据平面包括若干个和处理机连接的用于利用由控制路由器产生的报文上的通路信息发送和接收报文的数据路由器，以及一条用于把数据路由器和控制平面相同的拓扑结构连接在一起的数据线。

为达到第二个目的，提供一种在由一个控制平面和一个或多个具有和控制平面相同的拓扑结构的数据平面构成的互连网络中传送数据的方法，其包括步骤：从源节点向目的地节点沿着一条根据该拓扑结构的中间节点的通路传送带有目的地节点地址的标题信息，从而建立横跨控制平面传送报文的一条通路；在横跨控制平面的该通路上向数据平面传送信息，从而建立一条横跨数据平面用于传送报文的通路；释放该已建立的横跨控制平面的通路以便传送其他报文，并且沿着已建立的横跨数据平面的通路向目的地节点传送数据。

通过参照附图对本发明的最佳实施方式的详细说明本发明的上述目的和优点会更为清晰，附图是：

图1表示多计算机系统中互连网络和节点的总结构；

图2表示按照常规技术的网状平面结构；

图3对比按照常规技术的重叠互连网络情况和增大链路的带宽的情况；

图4A至4C表示按照常规技术网状平面中的报文传送过程；

图5表示根据本发明的具有一维结构的互连网络；

图6是一个流程图，表示根据本发明的数据传送方法；

图7表示作为本发明的一种实施方式的二维网状平面结构；

图8A至8C表示根据本发明的网状平面中的报文传送过程；以及

图9表示一种结构，其中根据本发明的互连网络中数据平面的带宽增至三倍。

下面参照附图详细说明本发明。

在本发明中,由物理上独立的网络分别运送在节点之间传送的数据和互连网络为传送数据所需的控制信号。传送控制信号的网络被称为控制平面,而传送数据的网络被称为数据平面。

如图5中所示,按照本发明的互连网络分离成控制平面和数据平面。控制平面通过接口适配器570、572和574和构成节点的处理机连接并且包括每个节点一个的控制路由器510、512和514,控制路由器用于交换设定报文通路的控制信息、管理互连网络和检测错误。控制平面还包括一条控制线540,用于按预定的拓扑结构连接控制路由器。数据平面通过接口适配器570、572和574和构成节点的处理机连接,并且包括每个节点一个的用于设定处理机间传送的报文通路的数据路由器520、522和524。数据平面还包括数据传送线550,用于按和控制平面相同的拓扑结构连接数据路由器520、522和524。

数据平面可扩充成大于1的预定数量的平面。图5表示可扩充的2号数据平面(包括数据路由器530、532和534以及数据传送线560),并且示出可扩充的数据平面的数量在接口适配器570、572和574所施加的限制之内。

图6是一个流程图,表示根据本发明的一种传送数据的方法。按照本发明,从源节点到目的地节点的通路是在控制平面中确定的(步骤600)。即,从源节点发送的标题包含目的地节点的地址。接收该标题的节点基于目的地节点的地址确定它的通路。当在控制平面上完成通路的确定时,该通路上的每个控制平面路由器向数据平面上的对应路由器给出通路信息(步骤610)。以这种方式在数据平面上构造通过它可以由源节点向目的地节点传送数据的通路。在完成设定通路的处理时,在控制平面上朝着目的地节点传送标题,并且在各数据平面(第一数据平面、第二数据平面、...)上数据随着标题水平地移动。因此,本发明称为Tornado路由方法(步骤620)。

本发明可应用于各种拓扑结构,例如,环、Banyan树、树、超立方体、网和环面。但是,在本发明的实施方式中,将说明二维网状平面结构的应用。

图 7 表示根据本发明的二维网状平面结构。在图 7 中所所示的二维网状平面结构中，细线代表控制平面而粗线代表数据平面。在图 7 中，数据平面只采用一个平面。但是，如图 9 中所示，物理上分离的数据平面是可扩充的。在本发明中，控制平面和数据平面使用不同的路由器，它们分别称为控制路由器和数据路由器。

图 8A 至 8C 表示在根据本发明的网状平面中传递报文的过程。在图 8A 中，来自 A 的报文已经正在沿着一一条设定的通路传送，而来自 B 和 C 的报文必须使用由 A 的报文占据的通路上的节点。在本发明中，因为数据平面是和控制平面分离的并且通路是在控制平面上设定的，B 和 C 的报文可以在不考虑报文 A 在数据平面上的路径下进行通路设定。一旦控制路由器在控制平面上设定通路，该通路信息被传送给和该相同节点对应的数据路由器。接着，利用该通路信息，数据路由器要以开始沿着由控制平面确定的通路传送报文。图 8B 表示一种情况，其中 B 和 C 的报文的通路已经设定，但 A 的报文仍阻塞这些通路。在 A 的通路释放之前 B 和 C 的报文可以传送到节点 (1, 1) 和 (2, 1)。从而，因为已经设定了 B 和 C 的通路，一旦 A 释放通路可以立即传送报文，没有为了设定 B 和 C 的通路造成的延迟，如图 8C 中所示。

图 9 表示一种为了把带宽增至三倍把数据平面的数量扩充二个的实施方式。请注意，该互连网络的带宽理论上是可扩展的。根据本发明，因为在这种互连网络里控制平面和数据平面是分离的有可能任意地扩展仅数据平面的带宽，并且因为设定通路时所消耗的等待时间与横跨该互连网络传送的数据的长度和频率无关有可能构造高速互连网络。因此，有可能构造一种利用宽的带宽和短的等待时间的适应性强的多计算机系统。

图1

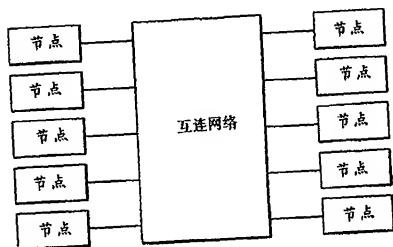


图2

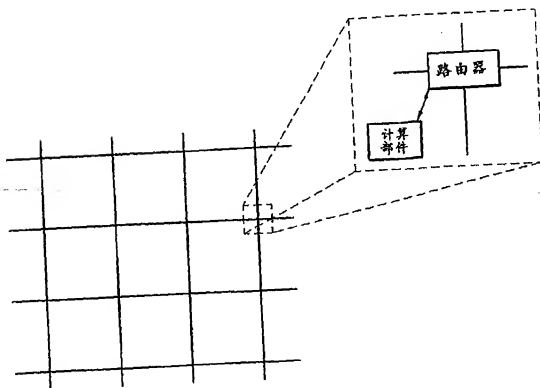


图 3

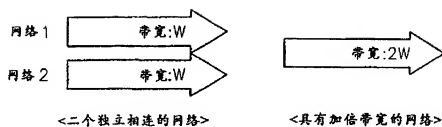


图 4A

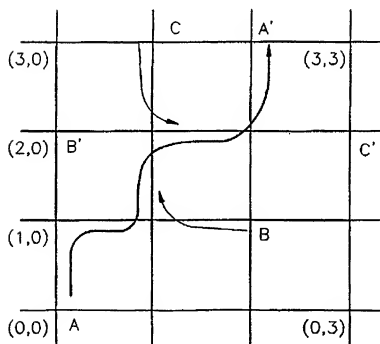


图 4B

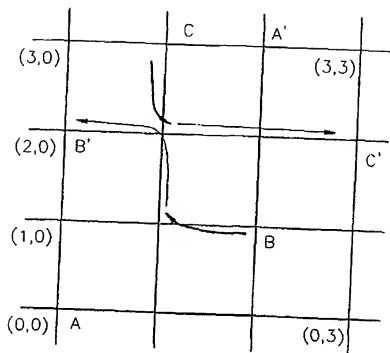
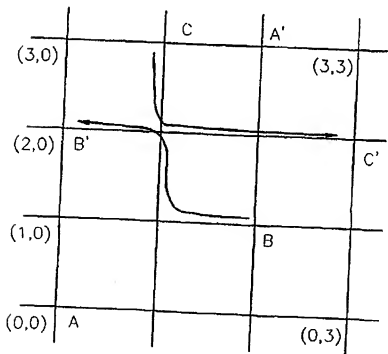


图 4C



可扩充的N
号数据平面

图 5

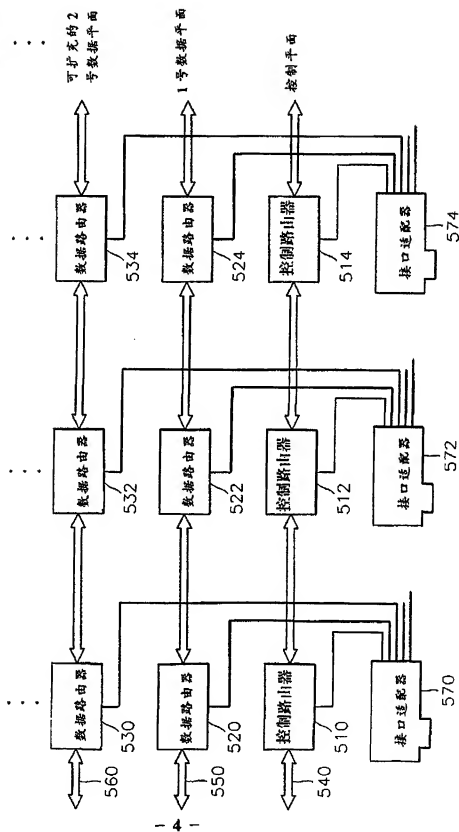


图 6

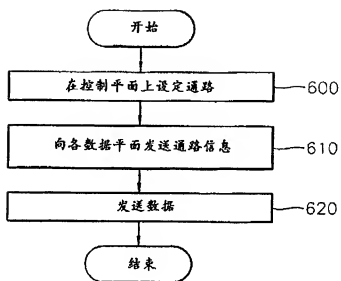


图 7

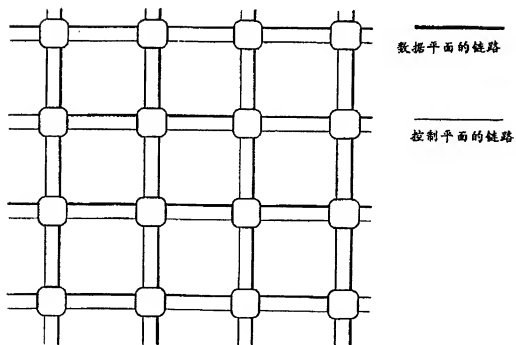


图 8A

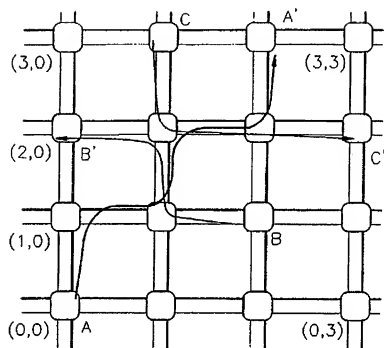


图 8B

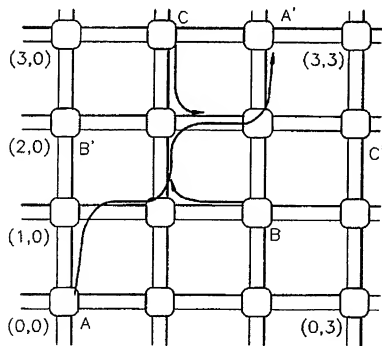


图 8C

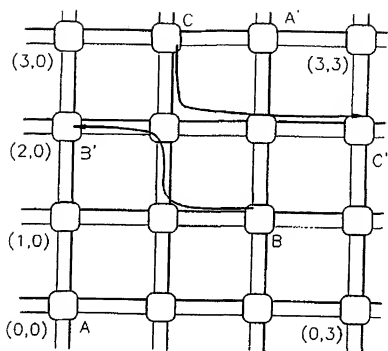


图 9

